

OSSベースでのビッグデータ分析を最大100倍に高速化 高速分析システムを低コストで構築可能に

日立は、FPGA※1上でオープンソースソフトウェア(OSS)のHadoopデータ処理基盤※2のデータベース処理を、並列実行する技術を開発しました。サンプルデータを用いて分析を行った結果、ビッグデータ分析を最大100倍高速化できることを確認。これにより、高速分析に必要なサーバ数を削減できるため、Hadoopベースのビッグデータ分析の低コスト化が可能です。日立は、今後お客さまとの実証実験を通じて、本技術の実用化をめざします。

※1 Field Programmable Gate Array: 製造後に購入者や設計者が構成を設定できる集積回路ハードウェア。特定用途向けの集積回路と比較して、低コストで開発が可能

※2 大規模データの配置や分析を、複数のサーバに分散させて処理するオープンソースのソフトウェアフレームワーク「Hadoop」を用いた計算基盤。サーバの単純な増加で処理性能を向上させることが可能

本技術の特長

近年大量かつ多様なデータを、さまざまな視点や条件でインタラクティブに分析して業務やサービスに反映させる、ビッグデータ活用の重要性が高まっています。その分析にはOSSのHadoopデータ処理基盤が広く使用されていますが、高速化のためには多数のサーバが必要で、機器や管理コストの増大が課題となっていました。

日立はこの課題を解決するため、本来ソフトウェア処理に合わせて作られているデータ処理手順をハードウェア上の並列処理に適した手順に変換したうえで、さまざまな形式のデータをFPGA上で並列に処理することにより、データベース処理を最大100倍に高速化する技術を開発しました。本技術を用いることで、高速分析のために多数のサーバを用意する必要がなくなり、高度なデータ分析を低コストで実現することが可能になります。

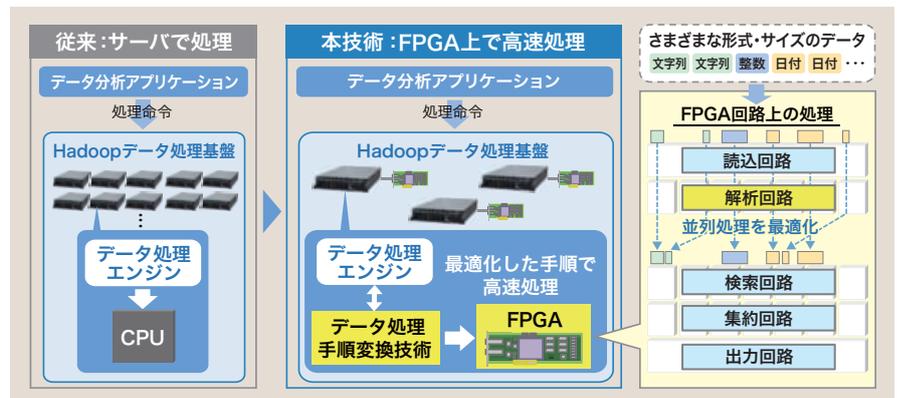
■必要サーバ数を削減可能

データ処理手順を、CPUによる順次処理に代わってFPGAによる並列処理にすることで、最大100倍の高速化を実現しました。同じ処理に必要なサーバ数を削減できるため、装置導入コストや消費電力、設置面積も低減できます。

■多様なデータを高速処理

FPGA上の処理では、従来はデータの種類ごとに専用の処理回路を用意する必要がありました。Hadoop処理基盤では複数の形式を許容するため、多くの専用処理回路が必要となり、搭載規模に限界のあるFPGAの回路領域を有効に活用できません。そこで、さまざまなデータ※3の形式やサイズを判別する解析回路を設け、ひとつの回路でまとめて処理するなど、並列処理の最適化調整を行う回路設計を考案しました。これにより、さまざまなデータに対応できるだけでなく、回路をむだなく利用した並列処理が可能となり、高速データ処理を実現します。

※3 Hadoopなど、オープンソースデータ処理基盤で一般的に用いられる標準データフォーマット「Parquet」をサポート



本技術の概要

適用例

■街中の混雑緩和／配車サービス

街中の車両や人の動きを位置データ分析に基づいて予測し、混雑緩和やタクシー配車などに活用。予測時間を短縮することにより、タイムリーなサービスが可能になります。

■金融商品レコメンド

お客さまの条件を聞きながら、金融機関の取引状況や市場情報をそのつど分析し、適切な金融商品を選んで提案。分析時間を短縮することで、対話中のお客さまをお待たせせずに商品を推薦することが可能になります。

お問い合わせ先・情報提供サイト

(株)日立製作所 研究開発グループ
<http://www.hitachi.co.jp/rd/>